

Development of a Bioinformatics Approach for the Functional Analysis of Alternative Splicing

Uno de los aspectos más apasionantes de la transcripción es la plasticidad transcriptómica y proteómica mediada por los procesos de regulación post-transcripcional (PTR). Los mecanismos PTR como el splicing alternativo (AS) y la poliadenilación alternativa (APA) han emergido como procesos estrechamente regulados que juegan un papel clave en la generación de la complejidad transcriptómica y están asociados con la coordinación de la diferenciación celular o el desarrollo de tejidos. Sin embargo nuestro conocimiento sobre cómo estos mecanismos regulan las propiedades de los productos resultantes para definir el fenotipo es aún muy reducido. La cantidad de variantes existentes y el amplio rango de posibles consecuencias funcionales, hacen su validación funcional una tarea impracticable si se realiza caso por caso. Además, la falta de herramientas para la evaluación funcional orientada a isoformas ha provocado que gran parte del trabajo computacional haya empleado pipelines ad-hoc aplicadas a sistemas biológicos específicos o simplemente hayan confiado en análisis de enriquecimiento GO, los cuales no son informativos del impacto en las propiedades de las isoformas que hay detrás de la regulación PTR.

De hecho, a pesar de las más de sesenta mil publicaciones relativas al AS, muy pocas isoformas se han asociado con propiedades específicas, mientras que el número de nuevas variantes AS/APA con funciones desconocidas e inexploradas crece exponencialmente gracias al empleo de las técnicas de secuenciación de segunda generación (NGS). Además, y debido a limitaciones técnicas de las NGS para reconstruir la estructura de los transcritos, las tecnologías de secuenciación de tercera generación (TGS) están definiendo una nueva era en la que, por primera vez, es posible conocer la secuencia de elementos estructurales y funcionales en una molécula de RNA.

En esta tesis se han abordado tres propósitos principales para poder avanzar en el estudio funcional de las isoformas. En primer lugar, con las TGS siendo cada vez más utilizadas, la definición y caracterización precisa de nuevos transcriptomas es esencial para asegurar la calidad en las conclusiones de la diversidad transcriptómica extraída de los datos. La falta de análisis de calidad orientados a secuencias largas ha motivado el desarrollo de SQANTI (<https://bitbucket.org/ConesaLab/sqanti>), una pipeline automatizada para la evaluación de la calidad de transcriptomas full-length. En segundo lugar, la información a nivel de gen de la mayoría de bases de datos funcionales sigue siendo el principal escollo para el estudio extensivo de la variabilidad de las isoformas, especialmente en el caso de las isoformas nuevas, en las que las bases de datos estáticas impiden su caracterización. Así, hemos diseñado IsoAnnot, que construye una base de datos de anotaciones funcionales con resolución a nivel de isoformas integrando información diseminada por múltiples bases de datos y métodos de predicción. Finalmente, la indisponibilidad de métodos para estudiar el impacto funcional de la regulación de isoformas, nos ha motivado a desarrollar tappAS (<http://www.tappas.org/>), una nueva herramienta dinámica, flexible, fácil de usar y diseñada para facilitar el abordaje de este tipo de estudios.

Por lo tanto, durante esta tesis hemos desarrollado una infraestructura que resuelve los retos principales del análisis funcional de isoformas, proporcionando un conjunto de nuevos métodos y herramientas que ofrecen una oportunidad única para explorar cómo el fenotipo se especifica post-transcripcionalmente, mediante la alteración de las propiedades funcionales de las isoformas expresadas. La aplicación de nuestro análisis a un doble sistema de diferenciación neuronal en ratón definió el efecto de la regulación de isoformas entre la diferenciación de motoneuronas y oligodendrocitos para múltiples elementos funcionales. Entre ellos, hemos descubierto regiones transmembrana que son diferencialmente incluidas en las isoformas expresadas entre ambos tipos celulares y cuya regulación podría estar contribuyendo al control de la dinámica mitocondrial, adaptando el metabolismo energético al tipo neuronal durante la diferenciación de células neuronales precursoras.